THE INVARIANT PROPERTY OF MAXIMUM LIKELIHOOD ESTIMATORS

ALLEN P. FANCHER

THE INVARIANT PROPERTY

OF

MAXIMUM LIKELIHOOD ESTIMATORS

* * * * * * *

Allen P. Fancher

THE INVARIANT PROPERTY

OF

MAXIMUM LIKELIHOOD ESTIMATORS

by

Allen P. Fancher

Lieutenant, United States Navy

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE

United States Naval Postgraduate School
Monterey, California

1 9 6 3

THE INVARIANT PROPERTY

OF

MAXIMUM LIKELIHOOD ESTIMATORS

by

Allen P. Fancher

This work is accepted as fulfilling

the thesis requirements for the degree of

MASTER OF SCIENCE

from the

United States Naval Postgraduate School

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

ABSTRACT

Classically, the invariant property of maximum likelihood estimators has been limited by one-to-one restrictions on the transformation. This thesis defines the Induced Likelihood Function and develops a theorem which may be used to extend the invariant property to estimation problems where the one-to-one restriction is dropped. It is shown that the theorem is applicable to the k dimensional estimation problem.

Theorem:

If 1) f is a function such that S is mapped into E_1 and $f(\hat{\Theta}) \geq f(\Theta)$ for all Θ in S.

2) ϕ is a transformation such that S is mapped into S^* where $\phi(\hat{\Theta}) = \Theta_o^*$ and $\phi(\Theta) = \Theta^*$ for all Θ in S. Define an inverse on S^* such that $\phi^{-1}(\Theta_o^*) = \hat{\Theta}$ and $\phi^{-1}(\Theta^*) = \Theta$ for all Θ^* in S^*.

3) g is a function defined by $g(\Theta^*) = f(\phi^{-1}(\Theta^*))$ such that S^* is mapped into E_1.

then $g(\Theta_o^*) \geq g(\Theta^*)$ for all Θ^* in S^*.

The writer wishes to express his appreciation to Professor P. W. Zehna for his guidance, assistance and the essential elements for the proof of the above theorem and to Professor J. R. Borsting for his suggestions and encouragement.

TABLE OF CONTENTS

SECTION I

INTRODUCTION

1.1 Statement of the Problem

The invariance property of maximum likelihood estimation provides a very convenient tool for statistical application. However, its use is somewhat limited in practical applications since the property apparently has only been shown to hold when the transformation of the parameter space is one-to-one. This investigation evolved as a follow-up to one phase of a reliability study undertaken by Captain W. J. Corcoran, USN, and Dr. H. Weingarten of the Technical Division of the Special Projects Office of the U. S. Navy and Dr. P. W. Zehna of CEIR (13)(14). This SP sponsored study presented several estimators for the parameters of a conceptual reliability model based on the multinomial probability distribution. One of the proposed estimates was "like" a maximum likelihood estimate (mle) in that it was a function of mle's; but since the function was not 1-1, the estimate was not formally called a mle. Attempts to derive distribution information concerning this estimate involved very complicated equations and these difficulties were compounded by the fact that under current definitions and concepts, maximum likelihood estimation (MLE) distribution theory was not correctly applicable. It was felt by Dr. Zehna that one of the primary questions that had to be answered prior to further work on the model was, "Does the invariance property of MLE apply when the functional relationship is not 1-1, and if so, under what conditions?"

1.2 Purpose of the Study

The purpose of this study is to investigate the invariance property of maximum likelihood estimation when the transformation function is not one-to-one, and to attempt to formalize concepts, definitions, and theorems that are applicable in this and similar situations.

1.3 Thesis Scope and Organization

Of necessity, it is assumed that the reader has a basic familiarity with the theory of probability and statistics, and the method and properties of maximum likelihood estimation. A brief review of some of the more pertinent concepts of point estimation along with a summary of the technique of MLE is presented in chapter two to provide a minimal common background and to assure familiarity with the notation as it is used in later discussions. Appendix one contains a chronological key to all notation used and is referenced by numbers indicating the page in the thesis on which the notation was originally introduced.

In chapter three the invariance property of MLE is discussed and concepts and theorems are developed which allow the present theory to be generalized and extended. Examples are liberally used to emphasize the points under discussion. Chapter four, in summary, attempts to indicate the possible contributions of the thesis, and suggests possible areas for further study.

SECTION II

MAXIMUM LIKELIHOOD ESTIMATION

2.1 Estimation; Basic Concepts

The purpose of statistical estimation is to estimate, on the basis of an observed sample, the values of the unknown parameters of the population from which the sample originated. Since 1763 when Bayes memoirs were published posthumously, there has been wide controversy and discussion concerning the various estimation techniques and the properties of the resulting estimates. Over the years several of these descriptive properties or characteristics have emerged as "desirable" traits of estimators. After presenting some concepts and definitions, several of the properties usually related to maximum likelihood estimates are briefly discussed.

Symbol/Term	Definition
$x_1, x_2, \ldots, x_n$	A sample or outcome of observed values of the random variables $X_1, X_2, \ldots, X_n$
$\theta_1, \theta_2, \ldots, \theta_n$	The parameters of an experiment - generally indices for some family of probability distributions
parameter	A constant of a probability distribution, generally unknown in estimation problems
$f(x;\theta)$	The probability density function of the random variable X with parameter indexed by θ, denoted pdf
$E(x)$	The expectation of x
estimator	A statistic; a rule for making an estimate of a parameter; a function of the observed values of the random variables. An estimator is derived <u>prior</u> to sampling.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]

Symbol/Term	Definition
estimate	A numerical value assigned to a parameter of a distribution on the basis of evidence from samples; an observed value of an estimator. An estimate is made after the sampling. (In this paper estimate will imply "statistical point estimate" unless otherwise indicated.)

The distinction between the parameter and its estimate is an important one. The true parameter value is fixed and unknown. However, with repetition of an experiment, the sample will vary, and the estimate itself will vary and will have a probability distribution. Estimation techniques are derived with the assumption that a sample is representative of the true population, therefore, the parameter estimate is subject to sampling errors. The possible magnitude of sampling error is an important consideration and leads to interval estimation which is not discussed in this paper.

2.2 The Method of Maximum Likelihood Estimation

The method of moments, introduced by Karl Pearson in 1894 was the earliest formal technique proposed for point estimation. Since that time many estimation procedures have been devised, the best known of which are the methods of minimum chi square, Bayes, Minmax, least squares, and maximum likelihood. It has been said that in many respects the introduction of maximum likelihood estimation marked the era of modern statistical theory.[1] The principle of maximum likelihood was discussed by Gauss prior to 1880, but R. A. Fisher formally developed maximum likelihood estimation (MLE) as a technique in a series of papers, the first of which was presented in 1921 (20).

[1] D. A. Fraser, Statistics, an Introduction, John Wiley and Sons, Inc., p. 224, 1958

Gauss had stated the concept in the following manner. Assume a random variable (vector) X with real values $(x_1, \ldots x_n)$ where the pdf of X is a function of the parameter(s) indexed by θ. Let θ have a known or assumed prior distribution with range a to b. Then the posteriori distribution of θ given X = x is

$$f(\theta|x) = \frac{f(\theta,x)}{\int_a^b f(\theta,x)d\theta} = c(x)f(\theta,x).$$

Gauss used the mode of the derived posteriori distribution as an estimate of θ. This value is what is commonly known today as the maximum likelihood estimate, the value of θ which maximizes the pdf of X with respect to θ.[2]

Fisher in his development derived what became known as the "Likelihood Function", the product of the population densities for each value in the sample. This function is denoted $L(\theta)$ where $L(\theta) = \prod f(x_i;\theta)$ and is regarded as a function of θ for fixed x_i. The method of maximum likelihood estimation is defined by maximizing this function. Since the logarithim is a monotonic increasing function, $L(\theta)$ and its log are maximized by the same value of θ. This is sometimes convenient since manipulation of log $L(\theta)$ is often much easier than working with the function directly.

The procedure for determining the mle of θ is as follows:

1) Determine the pdf, $f(x;\theta)$

2) Determine $L(\theta) = \prod f(x_i;\theta)$ and express as log $L(\theta)$ if appropriate.

3) Determine a value of θ which will maximize $L(\theta)$. This value is usually found by setting the derivative(s) of the likelihood function with respect to θ equal to zero and solving the ensuing equation(s) for the

[2]E. L. Lehmann, Notes on the Theory of Estimation, University of California, p. 1-9, 1950

parameter value(s) when conditions exist that make this possible. If $L(\theta)$ is differentiable and has its maximum at an interior point of the range of θ, the point at which $L(\theta)$ attains this maximum is the mle of θ, denoted $\hat{\theta}$, and the "Likelihood Equation" is $\left[\frac{\partial}{\partial \theta} \log L(\theta)\right]_{\theta=\hat{\theta}} = 0$.

Setting the derivative of a function equal to zero and solving in terms of a parameter does not in itself guarantee a maximizing value. If there is any doubt as to the authenticity of the solution, there should be further investigation to verify the underlying assumption, namely that the likelihood equations generally have only maximinizing solutions. Lindgren points out that this is usually the case since $L(\theta)$ is a product of probability densities and is usually bounded above and continuous in θ.[3]

2.3 Desirable Properties of Estimators

There are many ways that an estimator may be chosen. Hopefully statistical techniques provide the tools for choosing "good" estimators. To help describe what is meant by "good", several generally desirable properties or characteristics of estimators have been defined. The properties usually associated with mle's are discussed below.

1) Unbiasedness: This property is concerned with the distribution of the estimator. An estimator $\hat{\theta}(x_1, \ldots, x_n)$ for the parameter θ is said to be unbiased if $E(\hat{\theta}) = \theta$. Then, the bias of θ, denoted b, is $b = E(\hat{\theta}) - \theta$. Although unbiasedness is a desirable trait, it is by no means paramount. Figure 1 shows the densities of three estimators of θ. Although $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased, $\hat{\theta}_3$ is obviously the best estimator of the three even though it has positive or right bias. It is apparent that unbiasedness

[3] B. S. Lindgren, Statistical Theory, the Macmillian Company, p. 222, 1960.

considered alone does not guarantee a good estimator. The distribution, variance, and sample size all modify the bias.

2) Consistency: Consistency is a large sample property of an estimator. An estimator is said to be consistent if its probability distribution concentrates on the true parameter value as the sample size becomes infinite. That is, Θ is consistent if $P(|\Theta - \hat{\Theta}| < \delta) = 1$ as $n \to \infty$ for every $\delta > 0$.

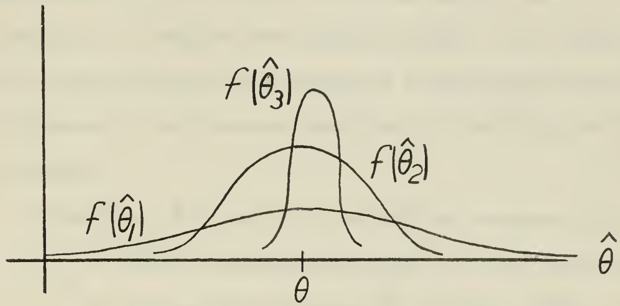

Figure 1. Density Functions of Three Estimators of the Parameter Θ[4]

An unbiased estimator is consistent if its variance approaches zero as the sample size approaches infinity.[5]

There may be many consistent estimators of a parameter. Therefore, as with unbiasedness, the criterion of consistency alone does not guarantee a useful estimator, although consistency is usually a desirable property.

3) Efficiency: Efficiency provides a criterion for comparing unbiased estimates of a parameter. As mentioned previously, once it is

[4]A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., p. 149, 1950.

[5]H. Cramer, Mathematical Methods of Statistics, Princeton University Press, p. 351, 1946

known that the distribution of the estimator is centered on the true value of the parameter, the variance of the distribution becomes an important consideration. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimates of θ, the "Relative Efficiency" of $\hat{\theta}_1$ to $\hat{\theta}_2$ is $(A_2/A_1) \times 100\%$ where $A_i = E(\hat{\theta}_i - \theta)^2$. If the ratio A_2/A_1 is greater than one, $\hat{\theta}_1$ may be considered a more efficient, and therefore perhaps a more suitable estimate than $\hat{\theta}_2$. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimates of θ, then A_2/A_1 is a ratio of variances and will take on its highest values when $\hat{\theta}_1$ is an estimate with minimum variance. R. A. Fisher proposed that the estimator having a minimum variance in large samples should be called "Efficient". This idea was formalized by a definition very similar to the following:

Definition: $\hat{\theta}$ is said to be an efficient estimator of θ if:

1) $\sqrt{N}(\hat{\theta} - \theta)$ approaches $N(0,\sigma^2)$ as N approaches infinity.

2) for any other estimator $\hat{\theta}^*$ for which $\sqrt{N}(\hat{\theta}^* - \theta)$ approaches $N(0,\sigma^{2*})$, $\sigma^{2*} \geq \sigma^2$. (The efficiency of θ^* is $(\sigma^2 / \sigma^{2*}) \times 100\%$.)

The Cramer-Rao inequality[6] may be used to find the limiting value of mean square deviations (variances for unbiased estimators). Efficient estimators are consistent[7] but are not necessarily unbiased except in the limit.

4) Sufficiency: An estimator is sufficient if, "it contains all the information in the sample regarding the parameter",[8] that is, it utilizes all of the pertinent information in the sample.

[6]ibid, p. 477

[7]A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., p. 151, 1950

[8]ibid

Definition: $\hat{\theta}$ is a sufficient estimator of θ if, given the value of $\hat{\theta}(x_1, \ldots, x_n)$, the conditional distribution is independent of the parameter θ.

In many situations the evaluation and manipulation of conditional distributions is very difficult, however, the following criterion allows determination of sufficiency by discerning if the joint density function can be properly factored.

Theorem 2.1: An estimator is sufficient if and only if the probability density function can be factored into two functions g and h, where h is dependent on the estimator and the parameter and g is independent of the parameter. That is, $\hat{\theta}$ is sufficient if $\prod f(x_i, \theta) = g(x_i, \ldots, x_n | \hat{\theta})\ h(\hat{\theta};\theta)$.

If sufficient statistics exist, it has been shown that they will be solutions of maximum likelihood.[9]

5) Invariance: This property, which is to be discussed at length in chapter three is usually associated with maximum likelihood estimation. The property implies that if the mle of θ is $\hat{\theta}$ and certain regularity conditions are satisfied a mle of $\phi(\theta)$ is $\phi(\hat{\theta})$. That is, a mle of a function of θ is simply the function with the value of $\hat{\theta}$ substituted for θ.

Maximum likelihood estimates are usually biased, consistent, efficient, invariant, and a function of a sufficient statistic if one exists. Under fairly general regularity conditions $\hat{\theta}$ is asymptotically normally distributed, has finite variance with limiting value = $1/I(\theta)$ where

[9] R. A. Fisher, Contributions to Mathematical Statistics, John Wiley and Sons, Inc., p. 224, 1958

$I(\theta) = nE\left\{\left[\frac{\partial}{\partial \theta} \log f(X,\theta)\right]^2\right\}$, and therefore is asymptotically Efficient.[10] No other Asymptotically normally distributed estimator can have smaller variance.[11] If an efficient statistic exists for small samples (i.e. with minimum variance), a mle with bias correction, if necessary, will be it.[12] This follows from the fact that if there is an unbiased efficient estimate, the maximum likelihood method will produce it.[13] Similarly, if there is a sufficient statistic for estimating the true parameter value, <u>any</u> solution of the likelihood equation will be a function of it.

From the preceeding summary, it can be seen why MLE has become a favored and often used technique in the field of statistical estimation. Although each of the estimation techniques has its strong points and proponents, (Pearson hotly defended the method of moments as "best" (44)), the mle is generally expected to exibit more of the desirable properties of a point estimator. Still, for certain instances, depending on the situation and problem at hand, the use of other estimation techniques may seem more logical and/or be easier. In fact, for certain distributions different techniques may produce the same estimate although generally they are different. The methods of moments and maximum likelihood produce the same estimates for the parameters of the normal, poisson, and binomial probability distributions.[14]

[10] H. Cramer, Mathematical Methods of Statistics, Princeton University Press, pp. 500-506, 1946

[11] A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., p. 160, 1950

[12] R. L. Anderson and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill Book Company, Inc., p. 102, 1952

[13] B. W. Lindgren, Statistical Theory, The Macmillian Company, p. 226, 1960

[14] S. S. Wilks, Mathematical Statistics, Princeton University Press, p. 146, 1943

2.4 Review of the Literature

A search of literature failed to yield any new and significant information concerning the application of the invariance principle to MLE. Of the college level statistics texts reviews, 17 contained sections on point estimation. Of these, only five discussed invariance as associated with maximum likelihood estimation, and eacn of these was restricted by the condition that the functional relationship be single valued or one-to-one. It is interesting to note that Mood[15] in his discussion of the property of invariance as applicable to MLE states that, " . . . if $\hat{\theta}$ is the maximum-likelihood estimate for θ, and if $u(\theta)$ is any single-valued function of θ, then $u(\hat{\theta})$ is the maximum-likelihood estimate for $u(\theta)$." However, in his proof of this property, it is implicitly assumed that an inverse function $\theta = v(u)$ is defined and he shows that the mle for u is the value of u that maximizes $L(v(u))$. Then, in addition to the necessity of having a single-valued function for the property to be applied as described by Mood, the inverse function must also exist. But even when the function is single-valued (but many to one, of course) there are many ways to define an inverse function. As shall be seen below, special care must be exercised in defining such an inverse. It also illustrates one of the situations motivating this investigation, namely, that discussions of the invariant property are often incomplete in the above sense.

A conspicious absence of literature concerning this property could be construed to indicate either that the problem is so trivial that it is unnecessary to record methods of application, or that the problem is of no practical or theoretical interest. Preliminary investigation of the problem

[15] A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., p. 159, 1952

[illegible]

at hand leads to rejection of both alternatives. The SP study mentioned in the opening pages of this paper is just one of many indicators that the problem is not trivial. Also, it provides a real practical application of the invariant property of MLE --- in an area not adequately covered by present concepts and definitions.

SECTION III

A NEW APPROACH

3.1 The Induced Likelihood Function

We have seen that previous applications of the invariance principle to the method of maximum likelihood estimation have been restricted so as to provide a 1-1 relationship between the domain and range of the functions of the parameters being estimated. Below a theorem is stated as it usually occurs in the 1-1 estimation problem.

Theorem 3.1:

If 1) $f: S \longrightarrow E_1$ (read, the function f map S into E_1)

2) ϕ: $S \xrightarrow[1-1]{onto} S^*$, therefore ϕ^{-1}: $S^* \xrightarrow[1-1]{onto} S$

3) g: $S^* \longrightarrow E_1$ defined by $g(\theta^*) = f(\phi^{-1}(\theta^*))$

4) there exists an element of S, denoted θ_o, such that $f(\theta_o) \geq f(\theta)$ for all θ in S

then $g(\theta_o^*) \geq g(\theta^*)$ for all θ^* in S^*

Proof: 1) Let θ^* be an element of S^*. Then $\phi^{-1}(\theta^*)$ is an element of S and

2) $g(\theta^*) = f(\phi^{-1}(\theta^*)) \leq f(\theta_o) = f(\phi^{-1}(\theta_o^*)) = g(\theta_o^*)$, i.e. $g(\theta_o^*) \geq g(\theta^*)$ for all θ^* in S^* and if θ_o is unique, strict inequality holds and θ_o^* is unique.

So far both the theorem and notation are conventional and application of the theorem to maximum likelihood estimation is as follows. Let S be the parameter space of the estimation problem. E_1 is the real line. The likelihood function $L(\theta)$ is such that L: $S \longrightarrow E_1$ and $L(\hat{\theta}) \geq L(\theta)$ for all θ an element of S. Suppose there exists a function ϕ such that ϕ: $S \xrightarrow[1-1]{onto} S^*$

so that $\phi^{-1}: S^* \xrightarrow[\text{1-1}]{\text{onto}} S$. Define the "Induced Likelihood Function, $M(\theta^*) = L(\phi^{-1}(\theta^*))$, the likelihood function induced on S^*. We now have the essential elements to apply theorem 3.1.

1) $L: S \longrightarrow E_1$

20 $\phi: S \xrightarrow[\text{1-1}]{\text{onto}} S^*$

3) $M: S^* \longrightarrow E_1$

4) $\hat{\theta}$ is the value of θ such $L(\hat{\theta}) \geq L(\theta)$ for all θ an element of S^*

Therefore if $\theta_o^* = \phi(\hat{\theta})$, by theorem 3.1 $M(\theta_o^*) \geq M(\theta^*)$ for all θ^* in S^* and a mle of θ_o^* is $\phi(\hat{\theta})$. If $\hat{\theta}$ is unique, then θ_o^* is unique.

Although it becomes apparent with application, let it be emphasized at this point that the concept of the induced likelihood function (ILF) and the manner in which it is defined is a most important element of the application of the theorem. A new likelihood function is defined on S^* and the mle is the parameter value in S^* which maximizes this new function.

Prior to looking at situation in which the 1-1 condition is dropped, consider the following interesting example which emphasizes the importance of the definition of the new likelihood function on the transformed parameter space S^*. Let all of the essential conditions of theorem 3.1 hold and let S^* be contained in S. The theorem still applies and along with conventional MLE procedures produces $\phi(\hat{\theta})$ as the mle of θ_o^*. That is, the MLE procedure on S is carried out as usual and produces $\hat{\theta}$, a mle of θ. However, in this case $L(\theta)$ is defined not only on S but on S^* as well. What happens when the likelihood function is restricted to S^*? Naturally, it is not expected that the restricted mle will always be the same as that produced in the unrestricted case since the unrestricted estimate may not be a member of S^*. However, the interesting fact is that $\phi(\hat{\theta})$ is not necessarily the restricted mle even when $\hat{\theta}$ is an element of S^*. As an example consider the exponential

distribution.

1) Let $f(x;\theta) = \theta e^{-\theta x}$ for $\theta > 0$

2) $L(\theta) = \prod f(x_i;\theta) = \theta^n e^{-n\bar{x}\theta}$

$$L'(\theta) = n\theta^{n-1}\, e^{-n\bar{x}\theta}\,(1-\bar{x}\theta)$$

$$\hat{\theta} = \frac{1}{\bar{x}}$$

3) Let $\lambda = \phi(\theta) = \frac{\theta}{1+\theta}$, so that $\theta = \phi^{-1}(\lambda) = \frac{\lambda}{1-\lambda}$

and $g(x;\lambda) = \left(\frac{\lambda}{1-\lambda}\right) e^{-\left(\frac{\lambda}{1-\lambda}\right)x}$ for $0 < \lambda < 1$

4) Let $M(\lambda) = L(\theta | 0 < \theta < 1)$ for the restricted estimation problem.

Then $\hat{\theta} = \frac{1}{\bar{x}}$ for $\bar{x} \geq 1$ and is undefined for $\bar{x} < 1$.

5) However, if $M(\lambda) = L(\phi^{-1}(\lambda)) = L(\frac{\lambda}{1-\lambda})$, all of the essential conditions of theorem 3.1 are fulfilled.

   1) $L(\theta)$: $S \longrightarrow E_1$
   2) $\phi(\theta)$: $S \xrightarrow[1-1]{onto} S^*$
   3) $M(\lambda)$: $S^* \longrightarrow E_1$ is defined by $M(\lambda) = L(\phi^{-1}(\lambda))$
   4) $\hat{\theta}$ is the value of θ such that $L(\hat{\theta}) \geq L(\theta)$ for all θ in S.

Therefore by theorem 3.1 $M(\lambda)$ is maximized by $\hat{\lambda} = \phi(\hat{\theta}) = \frac{\hat{\theta}}{1+\hat{\theta}} = \frac{1}{1+\bar{x}}$ and the restricted mle ($\frac{1}{\bar{x}}$ for $\bar{x} \geq 1$) is not equal $\phi(\hat{\theta}) = \frac{1}{1+\bar{x}}$. Had S^* not been contained in S, the defining of the ILF would be absolutely necessary since $L(\theta)$ would have no meaning on S^*.

Taking note of the use of the 1-1 property in conventional maximum likelihood estimation, it is seen that the assumption that ϕ is 1-1 is used only in defining $M(\theta^*)$ as a single valued function. If ϕ is not 1-1, how may the MLE problem be handled? As before, the key concept is the characterization of the new likelihood function and it can be shown that, with proper definition of the ILF, it is still maximized at $\phi(\hat{\theta})$.

Consider the case where $f : S \longrightarrow E_1$ and $\phi : S \xrightarrow{\text{onto}} S$, that is, the function is exaustive but not necessarily 1-1. The ILF, the likelihood function induced on S^* is defined in the following manner. If $L(\hat{\theta}) \geq L(\theta)$ for all θ an element of S, then let θ_o^* be any value of $\phi(\hat{\theta})$. Using the Axiom of Choice, if necessary, define an inverse on S^* such that $\phi^{-1}(\theta_o^*) = \hat{\theta}$ and for any other θ^* in S^*, $\phi^{-1}(\theta^*) = \theta$ where θ is any element of S such that $\phi(\theta) = \theta^*$. Then $\phi^{-1}: S^* \longrightarrow S$. Now theorem 3.1 can be extended and stated in a more general form.

Theorem 3.2:

If 1) $f: S \longrightarrow E_1$ and $f(\hat{\theta}) \geq f(\theta)$ for all θ in S

2) $\phi: S \xrightarrow{\text{onto}} S^*$, $\phi(\hat{\theta}) = \theta_o^*$

and $\phi^{-1}: S^* \longrightarrow S$ defined as above

3) $g: S^* \longrightarrow E_1$ defined by $g(\theta^*) = f(\phi^{-1}(\theta^*))$

then $g(\theta_o^*) \geq g(\theta^*)$ for all θ^* in S^*

Proof:

1) Let θ^* be an element of S^*

2) $g(\theta^*) = f(\phi^{-1}(\theta^*)) = f(\theta) \leq f(\hat{\theta}) = f(\phi^{-1}(\theta_o^*)) = g(\theta_o^*)$

Thus, $g(\theta_o^*) \geq g(\theta^*)$ for all θ^* an element of S^*

In the estimation problem let $M(\theta^*) = L(\phi^{-1}(\theta^*))$. Then $M(\theta_o^*) \geq M(\theta^*)$ so $M(\theta^*)$ is maximized by $\theta_o^* = \phi(\hat{\theta})$. The mle of θ_o^* is $\phi(\hat{\theta})$ just as in the 1-1 estimation situation. The maximization of $M(\theta^*)$ may not, in effect, have been over all the elements in S since ϕ^{-1} is not onto S, but it has taken place over the set containing $\hat{\theta}$ which is the essential factor.

Having repeatedly emphasized the importance of the definition of M, the ILF, it seems reasonable at this point to acknowledge the fact that

there may be many ways to define ϕ^{-1} and M. In some cases the definitions may be such that M is not maximized at $\phi(\hat{\theta})$ but this is not necessarily brought on by dropping the 1-1 restriction and in fact these same remarks apply even in the 1-1 case. In the restricted exponential estimation example, which was 1-1, two likelihood functions were defined on S^* and one was not maximized at $\phi(\hat{\theta})$.

Although the term "likelihood function" has been used extensively in theoretical statistics for quite a number of years, it appears that the term may be used rather loosely unless more emphasis is placed on the definition in a given problem. It is suggested that the notion of ILF may be an idea which will help to emphasize this point.

3.2 Examples of the Application of Theorem 3.2 and the Induced Likelihood Function to Maximum Likelihood Estimation

3.2.1 Geometric distribution

1) Let $f(x;\theta) = \theta(1-\theta)^{x-1} \qquad 0 \leq \theta \leq 1$

2) $L(\theta) = \theta^n(1-\theta)^{n(\bar{x}-1)}$

$$L'(\theta) = n\theta^{n-1}(1-\theta)^{n\bar{x}-n} - \theta^n\, n(\bar{x}-1)(1-\theta)^{n\bar{x}-n-1}$$

$$0 = n\hat{\theta}^{n-1}(1-\hat{\theta})^{n\bar{x}-n-1}\left[1-\hat{\theta}-(\bar{x}-1)\hat{\theta}\right]$$

$$\hat{\theta} = \frac{1}{\bar{x}}$$

3) Let $\lambda = \phi(\theta) = \begin{cases} \theta & \text{for } 0 \leq \theta \leq \frac{1}{2} \\ 1-\theta & \text{for } \frac{1}{2} \leq \theta \leq 1 \end{cases}$

$\therefore\ \phi : [0,1] \longrightarrow [0,\tfrac{1}{2}]$ and is not 1-1

4) Define $\phi^{-1}(\lambda) = \theta = \begin{cases} \lambda & \text{if } \bar{x} \geq 2 \\ 1-\lambda & \text{if } \bar{x} \leq 2 \end{cases}$

$\therefore\ \phi^{-1}: [0,\tfrac{1}{2}] \longrightarrow [0,1]$

5) Let $M(\lambda) = L(\phi^{-1}(\lambda)) = \begin{cases} L(\lambda) & \text{if } \bar{x} \geq 2 \\ L(1-\lambda) & \text{if } \bar{x} \leq 2 \end{cases}$

Therefore by theorem 3.2

$$\hat{\lambda} = \phi(\hat{\theta}) = \begin{cases} \hat{\theta} = \frac{1}{\bar{x}} & \text{if } \bar{x} \geq 2 \\ 1-\hat{\theta} = 1-\frac{1}{\bar{x}} & \text{if } \bar{x} \leq 2 \end{cases}$$

6) Checking the results directly

for $\bar{x} \geq 2$ $\quad M(\lambda) = L(\lambda)$ $\quad$ therefore $\hat{\lambda} = \frac{1}{\bar{x}}$

for $\bar{x} \leq 2$ $\quad M(\lambda) = L(1-\lambda)$

$$M(\lambda) = (1-\lambda)^n \left[1-(1-\lambda)\right]^{n(\bar{x}-1)}$$

$$M'(\lambda) = -n(1-\lambda)^{n-1} \quad \lambda^{n(\bar{x}-1)} + (1-\lambda)^n \, n(\bar{x}-1)^{n\bar{x}-n-1}$$

$$0 = n(1-\hat{\lambda})^{n-1} \, \hat{\lambda}^{n\bar{x}-n-1} \left[-\hat{\lambda} + (1-\hat{\lambda})(\bar{x}-1)\right]$$

$$1 = \bar{x}(1-\hat{\lambda})$$

$$\hat{\lambda} = 1-\frac{1}{\bar{x}}$$

3.2.2 Normal Distribution

1) Let $f(x;\theta) = \frac{}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$ for $-\infty < \theta < \infty$

2 $L(\theta) = (\frac{1}{2\pi})^{-\frac{n}{2}} \; e^{-\frac{1}{2}\sum(x_i - \theta)^2}$

$$L'(\theta) = (\frac{1}{2\pi})^{-\frac{n}{2}} \; e^{-\frac{1}{2}\sum(x_i - \theta)^2}\left[\sum(x_i - \theta)\right]$$

$$0 = \sum x_i - n\hat{\theta}$$

$$\hat{\theta} = \bar{x}$$

3) Let $\lambda = \phi(\theta) = \theta^2$

$\therefore \; \phi : E_1 \longrightarrow [0,\infty]$ and is not 1-1

4) Define $\phi^{-1}(\lambda) = \theta = \begin{cases} \sqrt{\lambda} & \text{if } \bar{x} \geq 0 \\ -\sqrt{\lambda} & \text{if } \bar{x} < 0 \end{cases}$

$\therefore \; \phi^{-1} : [0,\infty) \longrightarrow E_1$

5) Let $M(\lambda) = L(\phi^{-1}(\lambda)) = \begin{cases} L(\sqrt{\lambda}) & \text{if } \bar{x} \geq 0 \\ L(-\sqrt{\lambda}) & \text{if } \bar{x} < 0 \end{cases}$

Therefore by theorem 3.2

$\hat{\lambda} = \phi(\hat{\theta}) = \hat{\theta}^2 = \bar{x}^2$

6) Checking the results directly

if $\bar{x} \geq 0$, $M(\lambda) = L(\sqrt{\lambda}) = (\frac{1}{2\pi})^{-\frac{n}{2}} \quad e^{-\frac{1}{2}\sum(x_i - \sqrt{\lambda})^2}$

and $\sqrt{\hat{\lambda}} = \bar{x} \quad \therefore \; \hat{\lambda} = \bar{x}^2$

if $\bar{x} < 0$, $M(\lambda) = L(-\sqrt{\lambda}) = (\frac{1}{2\pi})^{-\frac{n}{2}} \quad e^{-\frac{1}{2}\sum(x_i + \sqrt{\lambda})^2}$

$$M'(\lambda) = -(\frac{1}{2\pi})^{-\frac{n}{2}} \quad e^{-\frac{1}{2}\sum(x_i + \sqrt{\lambda})^2} \left[\sum(x_i + \sqrt{\lambda})\right]$$

$$0 = \sum x_i = n\sqrt{\hat{\lambda}}$$

$$\sqrt{\hat{\lambda}} = -\bar{x} \qquad \therefore \; \hat{\lambda} = \bar{x}^2$$

3.2.3 Binomial Distribution

1) Let $f(x;\theta) = \theta^x(1-\theta)^{1-x} \quad x = 0, 1, \quad \text{for } 0 < \theta < 1$

2) $L(\theta) = \theta^{n\bar{x}}(1-\theta)^{n(1-\bar{x})}$

$$L'(\theta) = n\bar{x}\theta^{n\bar{x}-1}(1-\theta)^{n-n\bar{x}} - n(1-\bar{x})\theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}-1}$$

$$= n\theta^{n\bar{x}-1}(1-\theta)^{n-n\bar{x}-1}\left[\bar{x}(1-\theta) - \theta + \theta\bar{x}\right]$$

$$0 = \bar{x}(1-\hat{\theta}) - \hat{\theta} + \hat{\theta}\bar{x}$$

$$\hat{\theta} = \bar{x}$$

3) Let $\lambda = \phi(\theta) = \begin{cases} 2\theta \text{ for } 0 < \theta \leq \frac{1}{2} \\ 2 - 2\theta \text{ for } \frac{1}{2} < \theta < 1 \end{cases}$

$\therefore \quad \phi : (0, 1) \longrightarrow (0, 1)$ but is not 1-1

4) Define $\phi^{-1}(\lambda) = \theta = \begin{cases} \frac{\lambda}{2} & \text{if } \bar{x} \leq \frac{1}{2} \\ \frac{2-\lambda}{2} & \text{if } \bar{x} > \frac{1}{2} \end{cases}$

$\therefore \quad \phi^{-1} : (0, 1) \longrightarrow (0, 1)$

5) Let $M(\lambda) = L(\phi^{-1}(\lambda)) = \begin{cases} L(\frac{\lambda}{2}) & \text{if } \bar{x} \leq \frac{1}{2} \\ L(\frac{2-\lambda}{2}) & \text{if } \bar{x} > \frac{1}{2} \end{cases}$

Therefore by theorem 3.2

$$\hat{\lambda} = \phi(\hat{\theta}) = \begin{cases} 2\hat{\theta} = 2\bar{x} \text{ if } \bar{x} \leq \frac{1}{2} \\ 2 - 2\hat{\theta} = 2(1-\bar{x}) \text{ if } \bar{x} > \frac{1}{2} \end{cases}$$

6) Checking the results directly

If $\bar{x} \leq \frac{1}{2}$, $M(\lambda) = L(\frac{\lambda}{2}) = (\frac{\lambda}{2})^{n\bar{x}} (1 - \frac{\lambda}{2})^{n(1-\bar{x})}$

$$M'(\lambda) = n\bar{x}(\frac{\lambda}{2})^{n\bar{x}-1} (1 - \frac{\lambda}{2})^{n-n\bar{x}} \left[\bar{x}(1 - \frac{\lambda}{2}) - \frac{\lambda}{2}(1 - \bar{x})\right]$$

$$0 = \bar{x} - \frac{\hat{\lambda}}{2}$$

$$\hat{\lambda} = 2\bar{x}$$

If $\bar{x} > \frac{1}{2}$, $M(\lambda) = L(\frac{2-\lambda}{2}) = (\frac{2-\lambda}{2})^{n\bar{x}} \left[1 - (\frac{2-\lambda}{2})\right]^{n(1-\bar{x})}$

$$M'(\lambda) = n\bar{x}(\frac{2-\lambda}{2})^{n\bar{x}-1} \left[1 - (\frac{2-\lambda}{2})\right]^{n-n\bar{x}} \left[\bar{x}(1 - \frac{2-\lambda}{2}) - (\frac{2-\lambda}{2})(1-\bar{x})\right]$$

$$0 = \bar{x} - (\frac{2-\hat{\lambda}}{2})$$

$$\hat{\lambda} = 2 - 2\bar{x} = 2(1 - \bar{x})$$

3.3 The Multidimensional Estimation Problem

At this point, it seems logical to consider how the theorem applies in the estimation problem with a multidimensional parameter space. All examples considered to this point have been one-dimensional. However, since restrictions on dimensionality of the parameter space do not occur in the theorem or its proof, it follows that the theorem applies to the multidimensional estimation problem. Let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. If $\underline{\theta}$ is multidimensional, then so is $\underline{\hat{\theta}}$ and the components $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ are said to be the joint maximum likelihood estimates of the corresponding θ_i.

Consider the following example of the normal distribution with $\mu = \theta_1$, $\sigma^2 = \theta_2$ and $k = 2$.

1) $$f(x;\underline{\theta}) = f(x;\, \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \; e^{-\frac{1}{2}\frac{(x-\theta_1)^2}{\theta_2}}$$

2) $$L(\theta) = \left(\frac{1}{2\pi}\right)^{-\frac{n}{2}} (\theta_2)^{-\frac{n}{2}} \; e^{-\frac{1}{2}\sum\frac{(x_i-\theta_1)^2}{\theta_2}}$$

$$\frac{\partial L}{\partial \theta_1} = K\left(\sum x_i - n\hat{\theta}_1\right) = 0$$

$$\hat{\theta}_1 = \bar{x} = \hat{\mu}$$

$$\frac{\partial L}{\partial \theta_2} = -\frac{n}{\hat{\theta}_2} + \frac{1}{\hat{\theta}_2^2} \sum (x_i - \hat{\theta}_1)^2 = 0$$

$$\hat{\theta}_2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = S^2 = \hat{\sigma}^2$$

3) Let $\underline{\lambda} = (\lambda_1, \lambda_2) = \phi(\underline{\theta}) = \phi(\theta_1, \theta_2) = \left(\theta_1^2, \frac{\theta_2 - 1}{\theta_2}\right)$

ϕ is not 1-1

4) Define $\phi^{-1}(\underline{\lambda}) = (\theta_1, \theta_2) = \begin{cases} (\sqrt{\lambda_1}, \frac{1}{1-\lambda_2}) & \text{if } \bar{x} \geq 0 \\ (-\sqrt{\lambda_1}, \frac{1}{1-\lambda_2}) & \text{if } \bar{x} \leq 0 \end{cases}$

5) Define $M(\underline{\lambda}) = L(\phi^{-1}(\underline{\lambda})) = \begin{cases} L(\sqrt{\lambda_1}, \frac{1}{1-\lambda_2}) & \text{if } \bar{x} \geq 0 \\ L(-\sqrt{\lambda_1}, \frac{1}{1-\lambda_2}) & \text{if } \bar{x} \leq 0 \end{cases}$

Therefore by theorem 3.2

$$\underline{\hat{\lambda}} = \phi(\underline{\hat{\theta}}) = \phi(\hat{\theta}_1, \hat{\theta}_2) = (\hat{\theta}_1^2, \frac{\hat{\theta}_2-1}{\hat{\theta}_2})$$

6) Checking the results directly

if $\bar{x} \geq 0$, $\quad M(\underline{\lambda}) = L(\sqrt{\lambda_1}, \frac{1}{1-\lambda_2})$

$$= (\frac{1}{2\pi})^{-\frac{n}{2}} \ (\frac{1}{1-\lambda_2})^{-\frac{n}{2}} \ \exp -\frac{1}{2}\sum\frac{(x_i-\sqrt{\lambda_1})^2}{\frac{1}{1-\lambda_2}}$$

$$\frac{\partial M}{\partial \lambda_1} = K\left[\sum x_i - n\sqrt{\lambda_1}\ \right]$$

$$\hat{\lambda}_1 = \bar{x}^2 = \hat{\theta}_1^2$$

$$\frac{\partial M}{\partial \lambda_2} = \frac{\partial}{\partial \lambda_2}\left[K - \frac{n}{2} \log (\frac{1}{1-\lambda_2}) - \frac{1}{2}\sum(x_i - \sqrt{\lambda_1})^2 \ (\frac{1}{1-\lambda_2})^{-1}\right]$$

$$0 = \frac{-n}{(\frac{1}{1-\hat{\lambda}_2})} \quad + \sum \frac{(x_i-\sqrt{\hat{\lambda}_1})^2}{(\frac{1}{1-\hat{\lambda}_2})^2}$$

$$\frac{1}{1-\hat{\lambda}_2} = \frac{1}{n}\sum \ (x_i - \bar{x})^2 = S^2$$

$$\hat{\lambda}_2 = \frac{S^2-1}{S^2} = \frac{\hat{\theta}_2-1}{\hat{\theta}_2}$$

similarily if $\bar{x} \leq 0$

$$\underline{\hat{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2) = (\bar{x}^2, \frac{S^2-1}{S^2}) = (\hat{\theta}_1^2, \frac{\hat{\theta}_2-1}{\hat{\theta}_2})$$

In some situations we may desire to estimate only a portion of the θ_i. It should be noted that even though the estimates of only certain components of $\underline{\theta}$ are desired, it may be necessary to estimate the remaining parameters since the maximizing values for the desired set usually depend on the remaining parameters. This characteristic is demonstrated in the example just completed where the mle of the variance depends on the mle of the mean.

In estimating only certain of the components of $\underline{\theta}$ when the remaining parameters are unknown, theorem 3.2 is applied. However, if some of the remaining parameters are known, then the problem is quite different and the dimension of the parameter (estimation) space is reduced by one for each known parameter value. The problem of estimating the variance of a normal distribution with parameters $\mu = \theta_1$ and $\sigma^2 = \theta_2$ serves to illustrate this point.

Case I : μ , σ^2 unknown

We have seen that $\hat{\theta}_1 = \bar{x}$ and $\hat{\theta}_2 = S^2$. In this case, the parameter space is two-dimensional, a half-plane. That is $L : S \longrightarrow E_1$ where S is $E_1 \times (0,\infty)$.

Case II : μ known, σ^2 unknown

In this case $f(x;\underline{\theta}) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\theta_2}}$. Since μ is known, L is a function of θ_2 only and it is well known that $\hat{\theta}_2 = \frac{1}{n}\sum(x_i - \mu)^2$. In this case the problem is no longer to estimate a component of a two-dimensional $\underline{\theta}$, rather we have a new one-dimensional estimation problem where S is a subset of E_1.

Note that Case I produced $\hat{\theta}_2 = \frac{1}{n}\sum(x_i - \bar{x})^2$ and Case II produces $\hat{\theta}_2 = \frac{1}{n}\sum(x_i - \mu)^2$, usually quite different results. These differences, however, are not due to the application of the theorem, but result from the fact that the two likelihood functions are different.

SECTION IV

SUMMARY

4.1 Summary of Findings

The objective of this study was to investigate and formalize concepts and definitions that would allow the invariant property of MLE to be extended beyond the usually assumed 1-1 estimation situation. The induced likelihood function was introduced, and it has been shown that by properly defining the ILF, theorem 3.2 provides the tool for applying the invariance principle in the estimation problem with a transformation which is not 1-1. The theorem was shown to be equally applicable in the 1 or k dimension estimation situation.

In the development of theorem 3.2 it has been strongly emphasized that the power of the technique lies in the defining of the new likelihood function, the likelihood function induced on S^*. It is felt that, in the past, not enough emphasis has been focused on this induced likelihood function.

4.2 Proposed Areas for Further Study

This study has not attempted to investigate the distribution theory related to the mle's $\hat{\lambda} = \phi(\hat{\theta})$ derived using the ILF. Certainly, it is important to know if present mle distribution theory is still applicable in the unrestricted estimation situation. Therefore, it is suggested that an area which presents fertile ground for study is mle distribution theory in the new situations covered in this study.

The examples presented in this investigation are simple and are intended merely to acquaint the reader with the proposed use of the theorem and the ILF. It is hoped that this study has generated reader interest

which will result in application of the induced likelihood function and the associated theorem in a wide variety of estimation situations.

BIBLIOGRAPHY

1. J. Aitchison and S. D. Silvey, Maximum Likelihood Estimation Procedures and Associated Tests of Significances, J. Roy. Statist. Soc. Ser. B., 22, pp. 154-171, 1960.

2. R. L. Anderson and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill Book Company, Inc., 1952.

3. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., 1960.

4. R. R. Bahadur, Examples of Inconsistency of mle's, Sankhya, 20, pp. 207-210, 1958.

5. D. E. Barton, A Class of Distributions for which the mle is Unbiased and of Minimum Variance for All Sample Sizes, Biometrika, 43, pp. 200-202, 1956.

6. P. Billingsly, The Invariance Principle for Dependent Random Variables, Transactions of the American Mathematical Society, 83, p. 250, 1956.

7. D. Basu, An Inconsistency in the Method of Maximum Likelihood, Ann, Math Statist., 26, pp. 144-145, 1955.

8. A. Birnbaum, A Unified Theory of Estimation, Ann. Math. Statist., 32, p. 112, 1962.

9. D. Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions, John Wiley & Sons, Inc., 1954.

10. R. A. Bradley and J. J. Gart, The Asymptotic Properties of Maximum Likelihood Estimators When Sampling From Associated Populations, Biometrika, 49, pp. 205-214, 1962.

11. K. C. Chanda, A Note on the Consistency and Maxima of the Roots of the Likelihood Equations, Biometrika, 41, pp. 56-61, 1954.

12. H. Chernoff and L. E. Moses, Elementary Decision Theory, John Wiley & Sons, Inc., 1959.

13. W. J. Corcoran and H. Weingarten, Estimation of Reliability After Corrective Action on Observed Failure Modes, Special Projects Office, 1962.

14. W. J. Corcoran, H. Weingarten and P. W. Zehna, Estimating Reliability After Corrective Action, Unpublished Paper, 1962.

15. H. Cramer, Mathematical Methods of Statistics, Princeton University Press, 1946.

16. M. D. Donsker, An Invariance Principle for Certain Probability Limit Theorems, Memoirs Amer. Math. Soc. 6, p. 1, 1951.

17. J. Durbin and M. G. Kendall, The Geometry of Estimation, Biometrika, 38, pp. 150-158, 1951.

18. W. Feller, An Introduction to Probability Theory and Its Application, Vol. 1, John Wiley & Sons, Inc., 1950.

19. R. A. Fisher, Contributions to Mathematical Statistics, John Wiley & Sons Inc., 1950.

20. R. A. Fisher, On the Mathematical Foundations of Theoretical Statistics, Phil. Trans. Roy. Soc. Lonson Series A, Vol. 222, pp. 309-368, 1922.

21. R. A. Fisher, Theory of Statistical Estimation, Proc. Camb. Phil. Soc., Vol. 22, pp. 700-725, 1925b.

22. R. A. Fisher, Two New Properties of Mathematical Likelihood, Proc. Roy. Soc. Series, Vol. 144, pp. 285-307, 1934.

23. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley & Sons Inc., 1963.

24. D. A. S. Fraser, Statistics---An Introduction, John Wiley & Sons Inc., 1958.

25. D. A. S. Fraser, A Regression Analysis Using the Invariance Method, Ann. Math, Statist., 28, pp. 517-520, 1957.

26. D. A. S. Fraser, The Fiducial Method and Invariance, Biometrika, 48, pp. 261-280, 1961.

27. F. A. Graybill, An Introduction to Linear Statistical Models, Vol. 1, McGraw-Hill Book Company, Inc., 1961.

28. H. Hotelling, Experimental Determination of the Maximum of a Function, Ann. Math. Statis., 12, pp. 20-45, 1944.

29. S. Karlin and H. Rubin, Distributions Possessing a Monotone Likelihood Ratio, J. Amer. Statist. Assoc., 51, pp. 637-643, 1956.

30. M. G. Kendall and A. Stewart, The Advanced Theory of Statistics, Vol.1, Hafner Publishing Company, 1958.

31. M. G. Kendall and A. Stewart, The Advanced Theory of Statistics, Vol. II, Hafner Publishing Company, 1961.

32. J. Kiefer, Invariance, Minmax Sequential Estimation and Continuous Time Processes, Ann. Math. Statis., 28, pp. 573-601, 1957.

33. J. Kiefer and J. Wolfowitz, Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, Ann. Math. Statist., 27, pp. 887-906, 1957.

34. C. Kraft and L. LeCam, A Remark on the Roots of the Maximum Likelihood Equation, Ann. Math. Statist.

35. E. L. Lehmann, Notes on the Theory of Estimation, University of California, 1950.

36. E. L. Lehmann, Some Principles of the Theory of Testing Hypotheses, Ann. Math. Statist., 21, pp. 1-35, 1950.

37. B. W. Lindgren, Statistical Theory, The MacMillian Co., 1960.

38. D. E. Lloyd, A Note on a Problem of Estimation, Biometrika, 46, pp. 231-235, 1959.

39. D. K. Lloyd and M. Lipow, Reliability: Management, Methods, and Mathematics, Prentice-Hall, Inc., 1962.

40. A. M. Mood, Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., 1950.

41. J. Neyman, Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability, Phil. Trans. Roy. Soc., Vol. A 236, pp. 333-380, 1937.

42. J. Newman and E. L. Scott, Correction for Bias Introduced by a Transformation of Variables, Ann. Math. Statist., 31, pp. 643-655, 1960.

43. E. Parzen, Modern Probability Theory and Its Applications, John Wiley & Sons, Inc., 1960.

44. K. Pearson, Method of Moments and Method of Maximum Likelihood, Biometrika, 28, pp. 34-59, 1936.

45. R. B. Raja, A Formula for the Curvature of the Likelihood Surface of a Sample Drawn From a Distribution Admitting Sufficient Statistics, Biometrika, 47, pp. 203-207, 1960.

46. C. R. Rao, Maximum Likelihood Estimation for the Multinomial Distribution, Sankhya, 18, pp. 139-148, 1957.

47. F. S. G. Richards, A Method of MLE, J. Roy. Statis. Soc. Ser. B, 23, pp. 469-475, 1961.

48. K. Saito', Maximum Likelihood Estimate of Proportion Using Suplementary Information, Bull. Math. Statist., 7, pp. 11-17, 1956.

49. L. R. Shenton, Moment Estimators and Maximum Likelihood, Biometrika, 45, pp. 411-420, 1958.

50. S. D. Silvey, A Note on Maximum Likelihood in the Case of Dependent Random Variables, J. Roy. Statis. Soc. B, 23, pp. 444-452, 1961.

51. S. D. Silvey and T. W. Anderson, Maximum Likelihood Estimates for a Multivariate Normal Distribution When Some Observations are Missing, J. Amer. Statist. Assoc., 52, pp. 200-203, 1957.

52. S. D. Silvey and J. Aitchison, Maximum Likelihood Estimation of Parameters Subject to Restraints, Ann. Math. Statis., 29, pp. 813-828, 1958.

53. H. Steinhaus, The Problem of Estimation, Ann. Math. Statist., 28, pp. 633-648, 1957.

54. R. A. Rucker, An Introduction to Statistical Decision Functions, Thesis USN PGS, 1955.

55. C. Villegas, Maximum Likelihood Estimation of a Linear Functional Relationship, Ann. Math. Statist., 32, pp. 1048-1062, 1961.

56. A. Wald, Contributions to the Theory of Statistical Estimation and Testing Hypotheses, Ann. Math. Statist., 10, p. 299, 1939.

57. A. Wald, Note on the Consistency of the Maximum Likelihood Estimate, Ann. Math. Statist., 20, p. 595, 1949.

58. L. Weiss, Statistical Decision Theory, McGraw-Hill Book Company, Inc., 1961.

59. O. Wesler, Invariance Theory and a Modified Minmax Principle, Ann. Math. Statist., 30, pp. 1-20, 1959.

60. S. S. Wilks, Mathematical Statistics, Princeton University Press, 1943.

61. S. S. Wilks, Mathematical Statistics, John Wiley & Sons, Inc., 1962.

APPENDEX ONE

SYMBOLS AND ABBREVIATIONS

Symbol	Definition	Page*
mle	Maximum likelihood estimate	1
1-1	one-to-one	1
MLE	maximum likelihood estimation	1
x_i	observed value of random variable X_i	3
θ_i	index for i^{th} parameter	3
$f(x;\theta)$	probability density function	3
pdf	probability density function	3
$E(x)$	the expectation of x	3
$(x_1, x_2, \ldots, x_n)$	a sample or observed outcome	3
$f(\theta \mid x)$	the conditional pdf of θ given X=x	5
$L(\theta)$	the likelihood function	5
$\hat{\theta}(x_1, x_2, \ldots, x_n)$	an estimator	6
$f\colon S \longrightarrow E$	the function f is such that it maps S into E	13
E_1	the real line, Euclidean 1-space	13
$\phi\colon\ S \xrightarrow[1-1]{onto} S^*$	the function ϕ is such that it maps S onto S^* (onto implies "exaustive") and is 1-1.	13
M( )	the induced likelihood function	14
ILF	the induced likelihood function	14
$[0, 1]$	the closed interval 0, 1	17
$[0, 1)$	the half-closed interval 0, 1	18
$(0, 1)$	the open interval 0, 1	18
$\underline{\theta}$	$\theta_1, \theta_2, \ldots, \theta_k$	20

*Page on which symbol originally was introduced